Protein Profiling of Human Breast Tumor Cells Identifies Novel Biomarkers Associated with Molecular Subtypes*⑤

Anthony Gonçalves,[a,b,c,d] Emmanuelle Charafe-Jauffret,[c,e,f] François Bertucci,[b,c,f] Stéphane Audebert,[a] Yves Toiron,[a] Benjamin Esterni,[g] Florence Monville,[f] Carole Tarpin,[b] Jocelyne Jacquemier,[e,f] Gilles Houvenaeghel,[c,h] Christian Chabannon,[c,i] Jean-Marc Extra,[b] Patrice Viens,[b,c] Jean-Paul Borg,[a,c,j] and Daniel Birnbaum[f,j]

Molecular subtypes of breast cancer with relevant biological and clinical features have been defined recently, notably ERBB2-overexpressing, basal-like, and luminal-like subtypes. To investigate the ability of mass spectrometry-based proteomics technologies to analyze the molecular complexity of human breast cancer, we performed a SELDI-TOF MS-based protein profiling of human breast cell lines (BCLs). Triton-soluble proteins from 27 BCLs were incubated with ProteinChip arrays and subjected to SELDI analysis. Unsupervised global hierarchical clustering spontaneously discriminated two groups of BCLs corresponding to "luminal-like" cell lines and to "basal-like" cell lines, respectively. These groups of BCLs were also different in terms of estrogen receptor status as well as expression of epidermal growth factor receptor and other basal markers. Supervised analysis revealed various protein biomarkers with differential expression in basal-like *versus* luminal-like cell lines. We identified two of them as a carboxyl terminus-truncated form of ubiquitin and S100A9. In a small series of frozen human breast tumors, we confirmed that carboxyl terminus-truncated ubiquitin is observed in primary breast samples, and our results suggest its higher expression in luminal-like tumors. S100A9 up-regulation was found as part of the transcriptionally defined basal-like cluster in DNA microarrays analysis of human tumors. S100A9 association with basal subtypes as well as its poor prognosis value was demonstrated on a series of 547 tumor samples from early breast cancer deposited in a tissue microarray. Our study shows the potential of integrated genomics and proteomics profiling to improve molecular knowledge of complex tumor phenotypes and identify biomarkers with valuable diagnostic or prognostic values. *Molecular & Cellular Proteomics 7:1420–1433, 2008.*

Breast cancer (BC)[1] is a complex and heterogeneous disease resulting from accumulation of genetic alterations. This molecular heterogeneity explains in part the extensive diversity of clinical outcome and needs to be better delineated to improve therapeutic management and to identify relevant targets for novel treatments. A molecular taxonomy of BCs has been defined based on DNA microarray data (1–3). Five major molecular subtypes have been identified: luminal A and B, ERBB2-overexpressing, basal-like, and normal-like. These different BCs have a distinct clinical course and response to therapeutic agents (4–6). Overall luminal cancers (estrogen receptor (ER)-positive, 60% of BCs) have a good prognosis (although subtype B, which has a lower ER and higher proliferative profile, has a poor prognosis in comparison with subtype A). ERBB2-overexpressing (ER-negative and overexpressing ERBB2, 20–30% of BCs) and basal-like BCs (ER-negative and HER-2-negative, 10–20% of BCs) are unanimously considered as poor prognosis subtypes (1–3, 7). Importantly if molecularly targeted approaches are available for luminal (hormonal therapy) and ERBB2 BCs (trastuzumab), no similar treatment exists for basal-like BCs, justifying the need for a better molecular definition of this subtype. This definition may allow specific management and help identify novel molecular targets for innovative treatments.

Using whole-genome DNA microarrays and immunohistochemistry (IHC), we have recently established gene and protein expression profiles of 31 breast cell lines (BCLs) and generated a 10-protein signature (CAV1, CD44, EGFR, MET, ETS1, GATA3, luminal cytokeratin CK19, basal cytokeratin

[1] The abbreviations used are: BC, breast cancer; BCL, breast cell lines; ER, estrogen receptor; PR, progesterone receptor; IHC, immunohistochemistry; PTM, post-translational modification; TMA, tissue microarray; EGFR, epidermal growth factor receptor; bis-tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; IPI, International Protein Index; CMA, cell microarray; CI 95%, 95% confidence interval.

CK5/6, CD10, and Moesin) allowing accurate distinction between luminal and basal BCLs (8).

A promising way to complement molecular typing of tumors is to perform protein expression profiling using MS-based approaches (9). These approaches take advantage of the ability of mass spectrometers to separate peptides or proteins according to their *m/z*. They may identify peptides after enzymatic digestion of proteins separated from complex mixtures or may be applied directly to biological samples to generate a protein signature that correlates with a given phenotype. Theoretical advantages of this technology include the lack of requirement for an "*a priori*" hypothesis based on previous biological foreknowledge, allowing examination and quantification of a large number of initially unknown protein parameters as well as the potential to capture post-translational modifications (PTMs). PTMs are not detectable at the mRNA level but often play significant roles in protein functions. Among MS-based approaches, SELDI-TOF technology, which couples protein separation using chromatographic surfaces (ProteinChip arrays) and direct presentation to spectrometers, was made popular as a promising way to profile complex biological samples, notably biological fluids such as serum, plasma, or urine, to identify diagnostic or prognostic biomarkers (10–14).

Here we performed SELDI-TOF MS profiling of Triton-based protein lysates from 27 BCLs characterized previously at the transcriptional and IHC protein level. Our objectives were to explore how SELDI protein profiles may correlate with previously reported molecular subtypes identified in BCs and to identify specific biomarkers associated with these subtypes by taking advantage of specific MS features.

EXPERIMENTAL PROCEDURES

BCLs—27 of the 31 breast cell lines referenced in Charafe-Jauffret *et al.* (8) were analyzed: BT-20, BT-474, BT-483, CAMA-1, HCC1937, HCC38, HCC1954, HME-1, MCF-7, MCF-10A, MDA-MB-134, MDA-MB-157, MDA-MB-175, MDA-MB-231, MDA-MB-453, SK-BR-3, SK-BR-7, T47D, UACC-812, ZR-75-1, ZR-75-30, 184B5, BrCa-MZ-01, SUM-185, SUM-190, SUM-225, and S68. All cell lines are derived from human carcinomas except MCF-10A, which is derived from a fibrocystic disease, and HME-1 and 184B5, which represent immortalized normal mammary tissue. The cell lines were grown using the recommended culture conditions.

Protein Extraction—Cells were rinsed twice in cold PBS and lysed in buffer containing 50 mM HEPES, pH 7.5, 1 mM EGTA, 150 mM NaCl, 1.5 mM MgCl$_2$, 10% glycerol, 1% Triton X-100 supplemented with 1 mM PMSF, 1 mM orthovanadate, 10 nM aprotinin, and 1 μM leupeptin as antiprotease mixture. Triton-soluble proteins were recovered in the supernatant of a 20-min centrifugation at 13,000 × *g* and 4 °C. For tumors, frozen tissues were first cryoground and then subjected to the same lysis method.

Protein Expression Profiling—Protein concentrations were assessed using the Bradford assay, and an equal amount of total protein (20 μg) was investigated for each cell line. Samples were subjected to SELDI-TOF MS profiling using the ProteinChip Biomarker System as recommended by Ciphergen Biosystems (Fremont, CA). Briefly Triton-based cell lysates were bound in triplicate with a randomized chip/spot allocation scheme to IMAC-Cu and CM10

ProteinChip arrays. The energy absorbing molecule (crystallization matrix), 50% saturated sinapinic acid dissolved in 50% acetonitrile, 0.5% trifluoroacetic acid, was promptly applied. These steps were automated using a customized Tecan Evo Platform. All samples to be compared in a given experimental condition were processed in a one-step procedure. Spotted arrays were then read using a PBS IIC ProteinChip reader. For each experimental condition, readings were optimized for low molecular weight (2,000–30,000 Da). A pool of randomly spotted human serum specimens was used for monitoring the intra-assay reproducibility. External mass calibration was performed daily. Spectra were externally calibrated, base line-subtracted, and normalized to total ion current. Qualified mass peaks (signal/noise >5; cluster mass window at 0.3%) within the *m/z* range of 2–30 kDa were selected automatically using integrated Biomarker Wizard software. Resulting Excel files containing absolute intensity and *m/z* of protein peaks resolved were obtained and subjected to biostatistic processing.

Analysis of Proteomics Data—All data were log-transformed and analyzed by a combination of unsupervised and supervised methods. Unsupervised hierarchical clustering of expression data was done with the Cluster program (15) using Pearson correlation as the similarity metric and centroid linkage clustering. Results were displayed using the TreeView program (15). Supervised analysis was applied to the 327 peaks resolved and 27 cell lines to identify and rank proteins that discriminate between distinct relevant subgroups of cell lines using the nonparametric Wilcoxon Mann-Whitney test. Differentially expressed proteins were selected at an unadjusted *p* value of <0.05.

Protein Identification—Candidate biomarkers identified were then purified using IMAC-Cu-based chromatographic minicolumns (Hypercel, Ciphergen Biosystems, Fremont, CA) according to the manufacturer's instructions. These minicolumns allow recapitulating protein capture on IMAC ProteinChips as performed during the profiling phase. Briefly IMAC Hypercel columns were loaded with Cooper buffer and incubated with 300 μg of selected cell lysate samples in optimized binding buffer. After washing, proteins were eluted using 10 mM imidazole-containing buffer and concentrated to a final volume of 15 μl using a SpeedVac concentrator system. The purification process was monitored at all steps using NP20 ProteinChips. The purified biomarker was separated in an Xcell sure lock electrophoresis unit with 4–12% bis-tris gradient precast NuPAGE gels in MES running buffer according to the manufacturer's instruction (Invitrogen). Coomassie Blue-stained samples were washed, reduced, alkylated with 55 mM iodoacetamide, and digested at 37 °C for 16 h using 12.5 ng/μl specific enzymes (trypsin (Promega, Madison, WI) or endolysin (Sigma-Aldrich)) according to Shevchenko *et al.* (16). Peptides were extracted from the acrylamide gel by adding 75 μl of a 5% formic acid solution for 10 min and then 75 μl of the mixture acetonitrile/water/formic acid (60:35:5). Peptide extraction was increased using bath sonication. Extracted peptides were dried in a SpeedVac concentrator system and mixed with 4 μl of HCCA matrix solution (α-cyano-4-hydroxycinnamic acid in acetonitrile/water/trifluoroacetic acid (50:49.7:0.3)). 1 μl of the mixture was loaded on a standard Bruker 384 MALDI target plate. Mass spectrometry analyses were done with a MALDI-TOF instrument (Ultraflex, Bruker Daltonics, Billerica, MA) using reflectron and positive modes with an ion acceleration of 25 keV. 600 laser shots were accumulated for each spectrum. Mass spectra were processed with FlexAnalysis 2.0 software (Bruker Daltonics). Only peaks with a signal/noise higher than 5 were retained. Internal calibration with peptides 842.509, 1045.564, 2211.104, and 2283.180 corresponding to trypsin autolysis was used. A control spectrum corresponding to background peak (control piece of gel treated and digested the same as gel containing protein) was used to manually remove background peaks. Protein identification was carried out by peptide mass fingerprint using an in-house Mascot server (version 2.2.0), Matrix Science

Inc., London, UK. The MS spectra were searched against the International Protein Index (IPI) human database (version 3.26) from the European Bioinformatics Institute for peptide mass fingerprint identification. Criteria for searches were as follows: fixed carbamidomethylcysteine, optional methionine oxidation, no missing cleavage allowed, and a peptide search tolerance of 50 and 75 ppm for the trypsin and endolysin digest, respectively. Identification results were based on both the Mascot probability-based Mowse scores and the manual validation of mass assignments.

For Western blot analysis of full-length ubiquitin, cytosolic lysates from CAMA-1 and SUM-225 cells were separated by SDS-PAGE, transferred, and immunoblotted onto nitrocellulose as described previously (17) using anti-ubiquitin monoclonal antibody P4D1 (Cell Signaling Technology, Danvers, MA).

Gene Expression Profiling—RNA expression was profiled with Affymetrix U133 Plus 2.0 human oligonucleotide, representing over 47,000 transcripts and variants from human genes as described elsewhere (8).

Cell Microarray (CMA) and Tissue Microarray (TMA) Construction—The TMA from 547 patients with early breast cancer has been described previously (18). The CMA was constructed as described previously to circumvent the scattering of cells in paraffin-embedded cell lines (8). Briefly formaldehyde-fixed cell line pellets were resuspended at 37 °C in 1% low melting point agarose in 2-ml syringes and placed on ice, and the agarose cylinders obtained after cutting the terminal end of the syringe were fixed in ice cooled formalin-alcohol fixative. Cylinders were then processed in an automated tissue processor (ASP300, Leica) for an overnight run. The processed cylinders were then paraffin-embedded. CMA was prepared as for tissue microarrays with some modifications, mainly using a core cylinder with a diameter of 2 mm.

Immunohistochemistry—5-μm sections of the resulting blocks were made and used for IHC analysis after transfer onto glass slides as described previously (18) using a Dako LSAB^{R}2 Kit in the autoimmunostainer (Dako Autostainer, Glostrup, Denmark). Sections were deparaffinized in Histolemon (Carlo Erba Reagenti, Rodano, Italy) and rehydrated in a graded ethanol solution. Goat polyclonal anti-calgranulin B (C-19) antibody (Santa Cruz Biotechnology, Inc., Santa Cruz, CA) was applied at a dilution of 1:100. After staining, slides were evaluated by two pathologists (E. C. J. and J. J.). Results were scored by estimating the percentage (P) of tumor cells showing characteristic staining (from undetectable level or 0% to homogeneous staining or 100%) and by estimating the intensity (I) of staining (1, weak staining; 2, moderate staining; or 3, strong staining). Results were scored by multiplying the percentage of positive cells by the intensity, *i.e.* by the so-called quick score (Q) ($Q = P \times I$; maximum = 300). For each cell line and core biopsies, the mean of the score of a minimum of two core biopsies on two different slides was calculated. Discrepancies were resolved under the multiheaded microscope. For CMA, comparison between SELDI and IHC data were expressed as continuous values. For TMA, S100A9 staining had to be compared with other clinical and pathological data, and the cutoff value selected for S100A9 expression was >30 (median Q value of stained samples).

Statistical Analysis—Distributions of molecular markers and other categorical variables were compared using either the χ^2 or Fisher's exact tests. For continuous variables, Wilcoxon test was used. Metastasis-free survival was calculated from the date of diagnosis, the first distant metastasis being scored as an event. All other patients were censored at the time of the last follow-up, death, recurrence of local or regional disease, or development of a second primary cancer. Overall survival was calculated from the date of the diagnosis to the date of death or date of the last news. Survival curves were derived from Kaplan-Meier estimates (19) and compared by log-rank test. Survival rates are presented with their 95% confidence intervals (CI

95%). For multivariate analysis, Cox proportional hazards model regression was performed using a backward stepwise procedure based on Akaike information criterion. Statistical tests were two-sided at the 5% level of significance. All statistical tests were done using SAS version 8.02.

RESULTS

SELDI-based Protein Profiling of Breast Tumor Cell Lines and Correlation with Molecular Subtypes—Protein lysates from a total of 27 BCLs, previously characterized by IHC and DNA microarrays profiling (8), were profiled by SELDI-based mass spectrometry using CM10 and IMAC-Cu ProteinChip arrays. These two conditions generated a total of 326 protein peaks. Cell lines clustered according to the similarities of their SELDI-generated protein expression profiles, whereas proteins clustered according to their expression similarity across the sample population. Results of hierarchical clustering are shown in Fig. 1. BCLs displayed heterogeneous protein expression profiles as reflected by the dendrogram branch length. Overall they fell in two groups. Group I ($n = 12$) included only carcinoma cell lines with the majority (nine of 12) derived from ductal carcinoma (BT-483, UACC-812, T47D, ZR-75-1, MDA-MB-134, MCF-7, ZR-75-30, MDA-MB-175, and BT-474); the histological type of the other cell lines of this group (SUM-185, SUM-225, and S68) were not available. Group II ($n = 15$) comprised six ductal carcinoma cell lines (BT-20, HCC38, BRCA1-mutated HCC1937, HCC1954, SK-BR-7, and SK-BR-3), mesenchymal-like MDA-MB-231, two medullary BCLs (MDA-MB-157 and BrCa-MZ-01), three non-cancerous BCLs (184B5, MCF-10A, and HME-1), and three BCLs with unknown histological type (CAMA-1, MDA-MB-453, and SUM-190). A strong correlation existed between the two groups and the ER status of cell lines as determined in our previous study (8). This was true when ER status was evaluated either qualitatively by IHC ($p = 0.008473$, Fisher's exact test) with more ER-positive cell lines in group I (seven of 12) as compared with group II (one of 15) or quantitatively by DNA microarray-based measurement of *ESR1* gene expression ($p = 4.28 \cdot 10^{-5}$, Wilcoxon test).

We previously defined the same BCLs as "luminal-like" ($n = 13$) or "basal-like" ($n = 10$) according to breast cancer molecular subtyping generated from DNA microarray studies (2, 8, 20). As shown in Fig. 1, the major subgrouping of BCLs based on global clustering was in agreement with the subtype to which they were allocated: group I included 10 luminal-like cell lines of 13, and group II included nine basal-like cell lines of 10 ($p = 0.00275$, Fisher's exact test). This represents an 82% rate of concordance. Interestingly this SELDI-based subgrouping strongly correlated with differential expression of a molecular signature involving 10 potential basal markers (GATA3, CK19, EGFR, CD10, MET, CK5/6, CAV1, Moesin, CD44, and ETS1), which we generated from a DNA microarray study and validated by cell microarrays (supplemental Table 1). Thus, mass spectrometry-based profiling was able to capture protein expression information allowing the separation of
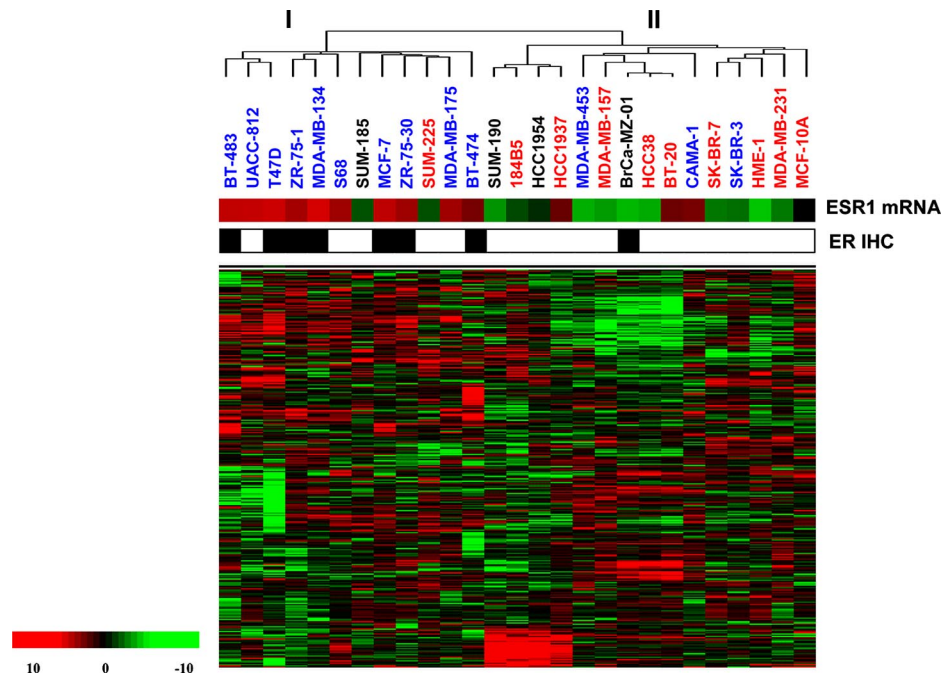
FIG. 1. **Global protein expression profiling of breast cell lines.** Hierarchical clustering of 27 mammary cell lines and 326 protein peaks based on SELDI measurements is shown. Each *row* represents a protein with a given *m/z*, and each *column* represents a cell line. The expression level of each protein in a single cell line is relative to its median abundance across all cell lines and is depicted according to a color scale shown at the *bottom*. *Red* and *green* indicate expression levels above and below the median, respectively. The magnitude of deviation from the median is represented by the *color saturation*. The dendrogram of samples (*above* matrix) represents overall similarities in protein expression profiles. Two groups of samples (designated *I* and *II*) are evidenced by clustering. The name of cell lines is colored as follows: *blue* for luminal-like (*n* = 13) and *red* for "basal /mesenchymal"-like (*n* = 10) cell lines according to previously described gene expression profiling (8). Four cell lines were not attributed to any subtype (name in *black*). The transcriptional (*ESR1*) and protein (*ER*) expression of ER of each cell line according to DNA microarray and IHC studies is represented.

FIG. 2. **Differentially expressed proteins according to molecular subtypes of breast cell lines.** *Top*, proteins with differential expression in luminal-like (*blue*) and basal-like (*red*) breast cells are plotted as a function of their normalized linear intensity (arbitrary units). Mean levels (*Mn*) in both molecular subtypes and *p* values (Mann-Whitney test) are indicated for each presented protein. *Bottom*, corresponding mass spectra in representative basal-like (*B*) and luminal-like (*L*) cell lines are shown in gel view, and differentially expressed protein are shown in *red circles*.
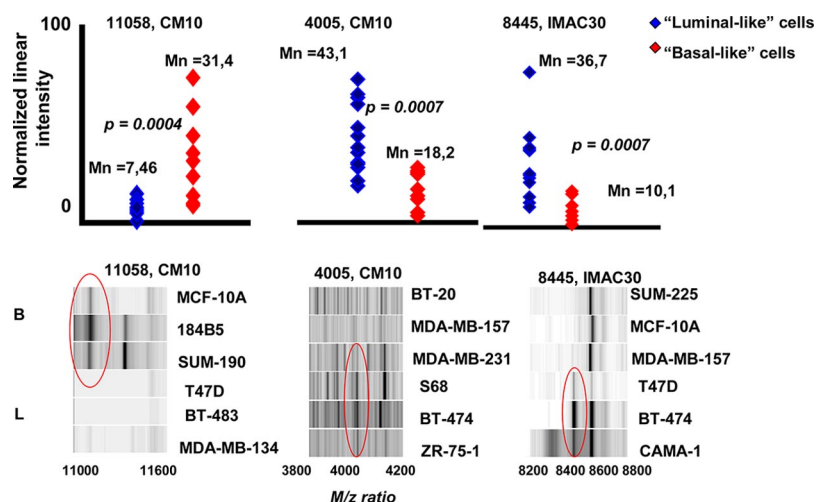
BCLs according to their major pathological and molecular features, including ER status and "luminal/basal" molecular subtyping.

Differential Protein Expression between Luminal- and Basal-like Subtypes—We then applied a supervised analysis based on the 326 protein peaks to the two breast subtypes defined above: luminal-like *versus* basal-like. We identified 73 protein peaks as potential discriminators with an unadjusted *p* value

threshold of 0.05, including 30 protein peaks that were overexpressed and 43 that were underexpressed in basal-like cell lines (supplemental Table 2). Proteins with the most significant differential expression between luminal- and basal-like BCLs are shown as histograms in Fig. 2. The hierarchical clustering of BCLs according to the expression of the 27 protein peaks with differential expression statistically significant at a *p* value less than 0.01 is illustrated in Fig. 3. The
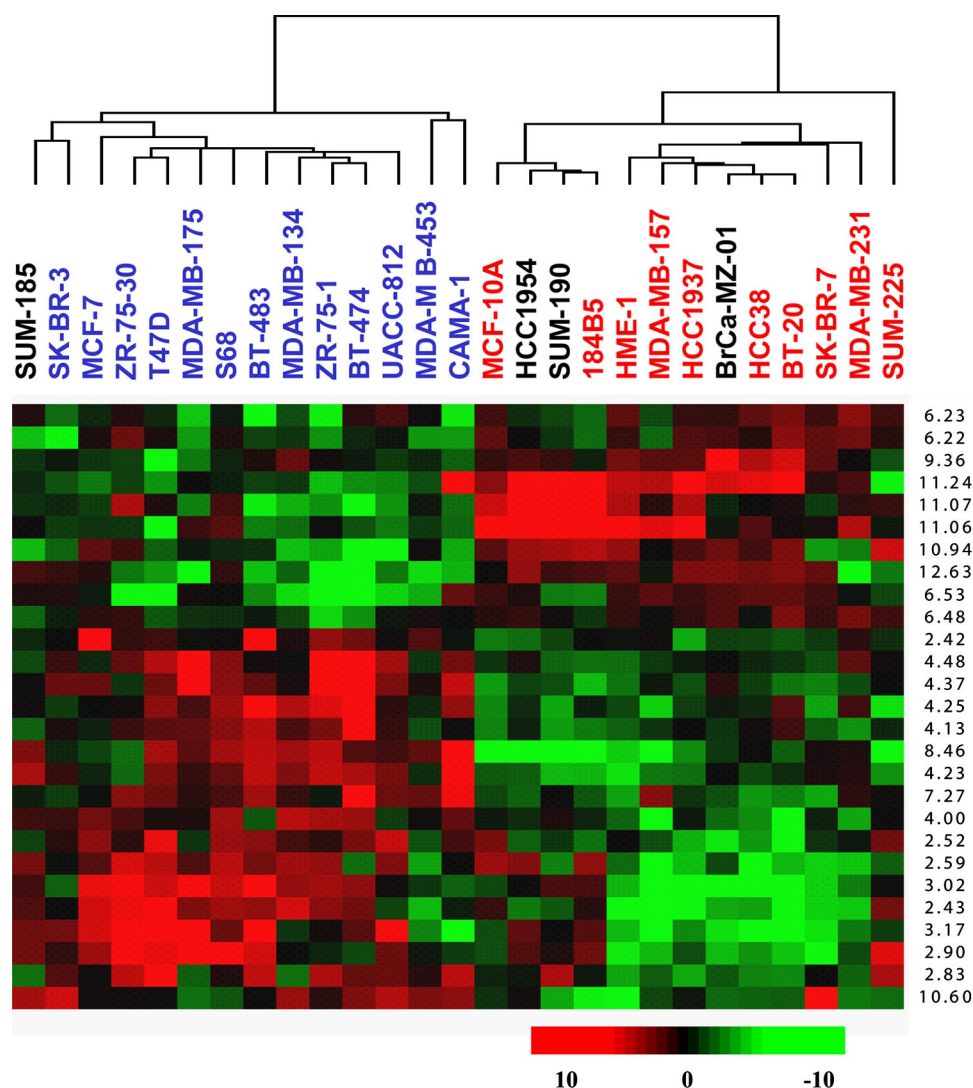
FIG. 3. **Supervised classification of 27 BCLs based on the 27 most significant differentially expressed protein peaks.** Each *row* of the data matrix represents a protein with a given *m/z* with differential expression between luminal-like and basal-like cell lines (*p* value <0.01), and each *column* represents a sample. Proteins and samples are ordered according to the similarity of their protein expression profiles (samples) or similarity of their expression across the sample population (proteins). *Color codes* are the same as in Fig. 1.

clustering was strongly correlated with the previously defined molecular subtyping. Thus, supervised analysis of SELDI-generated protein profiles for differential expression according to molecular subtyping revealed a large panel of candidate biomarkers.

Identification of a Truncated Isoform of Ubiquitin as a Potential Biomarker—A protein of *m/z* 8445, which was retained on IMAC-Cu ProteinChips, was very significantly down-modulated in basal-like cancer cells. Using two BCLs with drastic differential expression of this protein (CAMA-1 "up" and SUM-225 "down"; see Fig. 4*A*, *left*), the protein was subjected to purification and MALDI-based identification. 300 μg of protein lysates from CAMA-1 and SUM-225 were first incubated with IMAC Hypercell minicolumns, and then fractions containing the candidate biomarker were eluted using imidazole-based buffer. As seen in Fig. 4*A*, *left*, 8445 *m/z* protein was present

in the CAMA-1 sample but not in SUM-225 and was closely associated with another protein with *m/z* 8560 that was present in both samples. Eluted fractions were pooled and subjected to PAGE separation, and bands around 8 kDa were excised, subjected to trypsin digestion, and analyzed with MALDI-TOF MS for peptide mass fingerprint identification. Ubiquitin was unambiguously identified from the peptide mixture in both samples with very good (80%) sequence coverage (Fig. 4*B*, *left*, and Table I). The calculated mass of ubiquitin protein was in good agreement with the larger mass obtained experimentally with SELDI, *i.e.* a 8560 *m/z* peak. Surprisingly no supplementary protein corresponding to the 8445 *m/z* potential biomarker that was expected to be identified only in CAMA-1 could be evidenced. However, SELDI analysis of eluates from CAMA-1 purification as well as passive elution of protein bands cut and subjected to MALDI
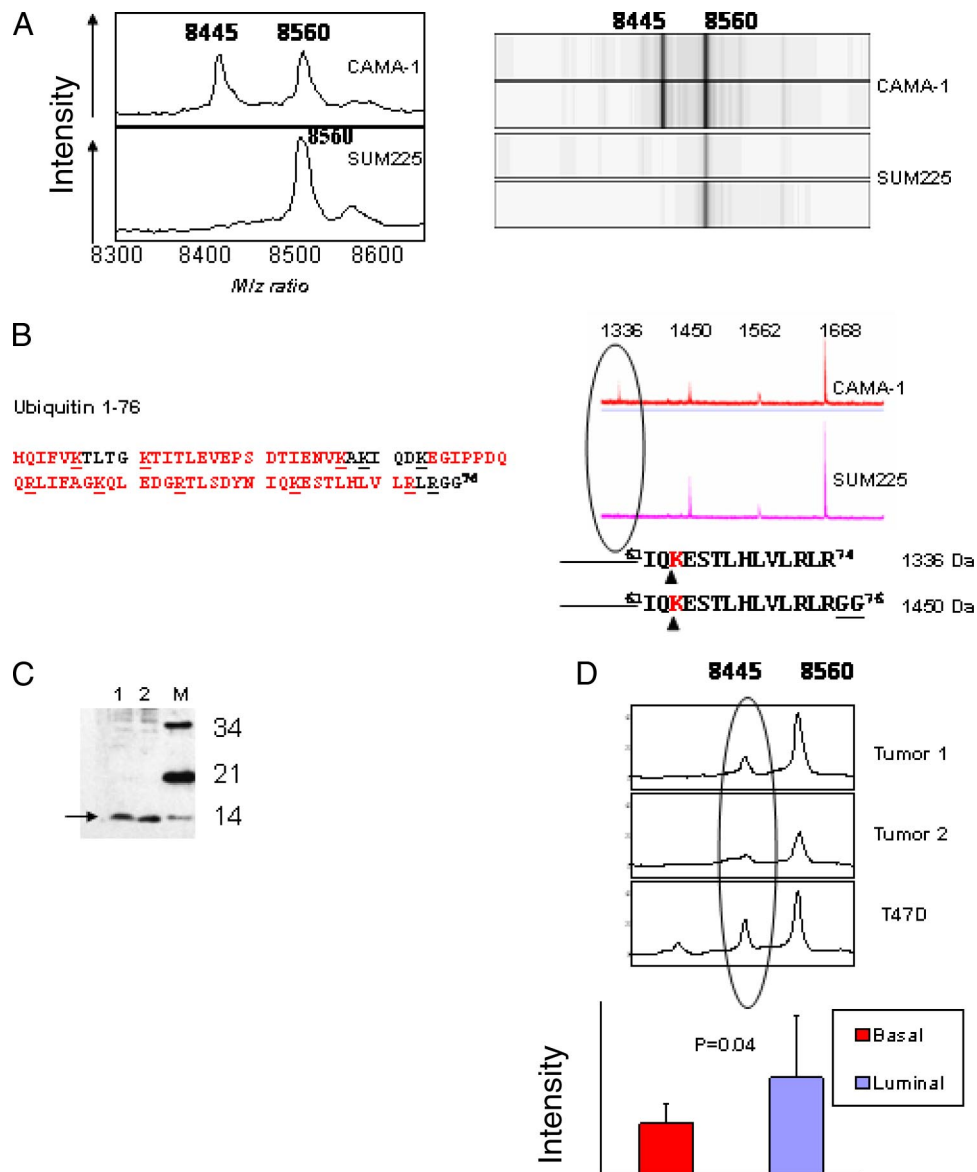
FIG. 4. **Purification and identification of a truncated form of ubiquitin as a potential luminal biomarker.** *A*, *left*, a potential biomarker of 8445 *m/z* retained on the IMAC30 ProteinChip array: 8445 is up-regulated in luminal-like CAMA-1 but not in basal-like SUM225 cancer cells (spectra view). *Right*, eluted fractions from IMAC30-based purification of 8445 protein in BCLs (gel view). Note that *m/z* 8560 (ubiquitin) is present in both CAMA-1 and SUM225 cells, whereas *m/z* 8445 is only present in CAMA-1cells. *B*, *left*, amino acid sequence of full-length ubiquitin. Sites of tryptic (Arg + Lys) or endolysin (Lys) are *underlined*. Coverage peptides identified by peptide mass fingerprint are shown in *red*. *Right*, endolysin digestion of CAMA-1 but not SUM-225 samples, generated an additional fragment of 1336 Da (*circle region*), corresponding to GG truncation of the ubiquitin carboxyl terminus. The intact carboxyl terminus was found in both samples as a 1450-Da peptide. Note that the *m/z* 1668 ion present in a relatively identical amount in both samples corresponds to the internal ubiquitin sequence (Glu^{34}–Lys^{48}). *C*, Western blot of BCLs lysates and anti-ubiquitin immunoblot. Levels of total ubiquitin were similar in both cell lines. *Lane 1*, CAMA-1; *lane 2*, SUM225; *lane M*, molecular weight markers. *D*, *top*, two frozen tumor tissues from early breast cancer were cryoground, subjected to Triton-based lysis buffer, and then incubated with the IMAC30 ProteinChip array. A control luminal-like cell line, T47D, was run in parallel. Truncated ubiquitin was found in breast cancer samples (*circle region*). *Bottom*, 11 frozen tissues from early breast cancer samples (including the two from the *upper panel*) that had been characterized by transcriptional analysis as luminal-like ($n = 6$) or basal-like ($n = 5$) were subjected to SELDI. Intensities of the *m/z* 8445 peak were measured, and mean intensities in both cell types were compared (Wilcoxon Mann-Whitney test). Truncated ubiquitin was significantly higher in luminal samples. *Error bars* represent standard deviations.

revealed that the 8445 *m/z* peak was actually present in CAMA-1 samples (Fig. 4*A*, *right*). This apparent discrepancy raised the possibility that the 8445 *m/z* potential biomarker was a post-translational modification of the same protein, *i.e.*

ubiquitin. Indeed this peak could be very reasonably be explained as a loss of two carboxyl-terminal glycines from ubiquitin (theoretical mass of cleaved ubiquitin = 8445, identical to experimental data). Because expected amino acid sites of

TABLE I

Ubiquitin identification by peptide mass fingerprint after trypsic digest of the protein bands

Ubiquitin was unambiguously identified in both CAMA-1 and SUM-255 cells line with 72% sequence coverage. Experimental (expt) masses matching calculated (calc) ubiquitin peptides are shown in red (results from SUM-225 extract are shown).

Sample	Protein Name	Accession Number	Mascot Score	Matched peptides	Unmatched peptides	Protein coverage (%)
Cama-1	Ubiquitin	IPI 00783060	115	6	16	72
Sum-225	Ubiquitin	IPI 00783060	116	6	15	72

Ubiquitin sequence coverage

```
1   MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL
51  EDGRTLSDYN IQKESTLHLV LRLRGG
```

Matching peptides
```
Start - End   Observed    Mr(expt)    Mr(calc)     ppm    Miss Sequence
    1 -  6     781.4190    780.4118    780.4204     -11     0  -.MQIFVK.T  Oxidation (M)
   12 - 27    1787.8793   1786.8720   1786.9200     -27     0  K.TITLEVEPSDTIENVK.A
   34 - 42    1039.5151   1038.5079   1038.5094      -1     0  K.EGIPPDQQR.L
   49 - 54     717.3639    716.3566    716.3453      16     0  K.QLEDGR.T
   55 - 63    1081.5438   1080.5365   1080.5451      -8     0  R.TLSDYNIQK.E
   64 - 72    1067.6161   1066.6088   1066.6135      -4     0  K.ESTLHLVLR.L
```

TABLE II

Ubiquitin-containing gel pieces from CAMA-1 and SUM-225 cells lines were digested with endolysin that cleaves specifically after lysine residues

Ubiquitin was identified with 84% sequence coverage. Results of Mascot peptide mass fingerprint corresponding to matched peptides are shown below. Note that the last amino acids LRGG are detectable in carboxyl-terminal peptides after endolysin cleavage (this table) but not after trypsin cleavage (Table I). expt, experimental; calc, calculated.

Sample	Protein Name	Accession Number	Mascot Score	Matched peptides	Unmatched peptides	Protein coverage (%)
Cama-1	Ubiquitin	IPI 00783060	111	5	14	84
Sum-225	Ubiquitin	IPI 00783060	119	5	10	84

Mascot Results :
Peptides matching to human ubiquitin sequence are shown in red

```
1   MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL
51  EDGRTLSDYN IQKESTLHLV LRLRGG

Start - End    Observed    Mr(expt)    Mr(calc)     ppm    Miss   Sequence
    2 -  6     634.3798    633.3725    633.3850     -20     0     M.QIFVK.T
   12 - 27    1787.8506   1786.8434   1786.9200     -43     0     K.TITLEVEPSDTIENVK.A
   34 - 48    1668.8439   1667.8366   1667.8995     -38     0     K.EGIPPDQQRLIFAGK.Q
   49 - 63    1779.8146   1778.8073   1778.8799     -41     0     K.QLEDGRTLSDYNIQK.E
   64 - 76    1450.7992   1449.7919   1449.8416     -34     0     K.ESTLHLVLRLRGG.-
```

tryptic digestion include arginine, such a truncation would not be detectable by MALDI analysis of trypsin-digested fragments (see Fig. 4*B*, *left*, the amino acid sequence of ubiquitin and theoretical sites of tryptic digestion, Lys and Arg *underlined*). To examine this hypothesis, we subjected CAMA-1 and SUM-225 bands to differential digestion using endolysin-C, which does not cut after both lysine and arginine but only after lysine (Fig. 4*B*, *left*). As illustrated in Fig. 4*B*, *right*, endolysin-C digestion of CAMA-1 but not SUM-225 samples generated an additional fragment of 1336 *m/z*, the sequence of which appeared to be deleted of two carboxyl-terminal glycines. The *m/z* ion of 1450 was present in both cases and corresponded to the calculated masse of the unmodified carboxyl-terminal end (Table II). Thus, a carboxyl terminus-truncated form of ubiquitin was the actual biomarker, whereas total ubiquitin was not different between the two cell types as determined by Western blot (Fig. 4*C*). We also examined the mRNA expression of cDNA sequences coding for ubiquitin proteins in the panel of BCLs studied. For ubiquitin-coding genes present on the array and capable of being evaluated for analysis, we found no difference in gene expression between luminal-like and basal-like BCLs (data not shown). We also looked at the gene signature associated with basal/luminal discrimination in a panel of human breast cancer analyzed by DNA microarrays (see below). No ubiquitin-coding sequence was found to be significantly differentially expressed according to the basal or luminal subtypes (data not shown). Thus, MS-based profiling can detect molecular variations that may not be identified by other high throughput technologies.

To evaluate the potential clinical relevance of this PTM in breast cancer, we analyzed two frozen tumor tissues by

TABLE III
S100A9 identification

Mascot search results against the IPI (European Bioinformatics Institute) are shown. Matched peptides are indicated in red on the protein sequence as well as mass/peptide assignments below. expt, experimental; calc, calculated.

Sample	Protein Name	Accession Number	Mascot Score	Matched peptides	Unmatched peptides	Protein coverage (%)
BT-474	S100-A9	IPI 00027462	144	7	23	80

S100-A9 sequence

```
1    MTCKMSQLER NIETIINTFH QYSVKLGHPD TLNQGEFKEL VRKDLQNFLK
51   KENKNEKVIE HIMEDLDTNA DKQLSFEEFI MLMARLTWAS HEKMHEGDEG
101  PGHHHKPGLG EGTP
```

```
Start - End    Observed    Mr(expt)    Mr(calc)       ppm   Miss Sequence
   11 - 25     1806.8964   1805.8891   1805.9312      -23      0  R.NIETIINTFHQYSVK.L
   26 - 38     1455.6933   1454.6860   1454.7154      -20      0  K.LGHPDTLNQGEFK.E
   44 - 50      877.4756    876.4683    876.4705       -3      0  K.DLQNFLK.K
   58 - 72     1742.7830   1741.7757   1741.8192      -25      0  K.VIEHIMEDLDTNADK.Q
   73 - 85     1614.7682   1613.7609   1613.7945      -21      0  K.QLSFEEFIMLMAR.L
   86 - 93      971.4840    970.4767    970.4872      -11      0  R.LTWASHEK.M
   94 - 114    2175.9425   2174.9352   2174.9552       -9      0  K.MHEGDEGPGHHHKPGLGEGTP.-
```

SELDI. As shown in Fig. 4*D*, the same 8445 *m/z* peak was observed. Frozen tissues were then subjected to the same purification procedure on IMAC minicolumns as above. After PAGE separation and excision of the appropriate bands, MALDI identification of the same carboxyl terminus-truncated fragment was demonstrated after endolysin-C digestion (data not shown), indicating that such a PTM occurs also in clinical tumor samples. We analyzed a total of 11 frozen whole breast cancer samples, which had been analyzed previously by DNA microarray and classified as basal or luminal. As shown in Fig. 4*D* (*lower panel*), the truncated form of ubiquitin had a significantly higher expression in luminal breast cancer samples (mean intensity of 197 ± 134 arbitrary units) as compared with basal breast cancer samples (mean intensity of 104 ± 43 arbitrary units) ($p = 0.04$, Wilcoxon Mann-Whitney test), although there was no significant difference in full-length ubiquitin expression (data not shown).

Identification of S100A9 as a Potential Biomarker and Validation on Clinical Samples by DNA Microarray- and TMA-based Approaches—Using a similar purification and MALDI-based identification process (data not shown), a protein peak of *m/z* 13,300, with differential expression according to the molecular subtype, was retained on CM10 ProteinChips and was identified as S100A9 protein (also called calgranulin B or MRP-14; Table III). To validate the clinical relevance of this potential biomarker of basal subtype, we evaluated S100A9 expression on tumor samples at both the transcriptional and protein levels by using different expression profiling techniques.

First, we examined S100A9 mRNA expression in a panel of 148 of early breast cancers and four normal breast samples that we had profiled previously using whole-genome Affymetrix DNA microarrays. Hierarchical clustering applied to the 13,743 genes/expressed sequence tags with significant variation in expression level across all samples is shown in Fig. 5*A*. In this panel, tumor samples had been characterized for their relationship to molecular subtypes using the Stanford intrinsic 500-gene set (2) as described previously (21). As expected, luminal samples and basal samples clustered in two separate major groups (Fig. 5*B*). Several clusters of related genes were evidenced, corresponding to specific cell types or pathways (see *colored bars* to the *right* of Fig. 5*A*). Among them, the luminal and the basal clusters had a prominent role in the classification of samples. Interestingly S100A9, which very closely clustered with S100A7 and S100A8, was included in the basal cluster (Fig. 5*B*) in proximity with several cytokeratins (*KRT5*, *-6*, *-14*, *-15*, and *-17*), *EGFR*, *CRYAB*, and *CDH3*.

Second, we measured S100A9 protein expression by IHC. To validate the selected antibody, we analyzed S100A9 expression by IHC on a CMA containing 25 BCLs used in the MS profiling phase. As shown in Fig. 6*A*, IHC and SELDI data were significantly correlated (correlation coefficient = 0. 42; *p* value = 0.045, Pearson test). Then S100A9 expression was evaluated by IHC on a TMA containing 1600 specimens from 547 patients with early BC (18). Staining was available for 386 patients and considered positive when located in the cytoplasm of tumoral cells. A total of 115 (29%) samples expressed S100A9 (Fig. 6*B*). Correlations of S100A9 expression with clinical, pathological, and molecular features of the population were explored (supplemental Table 3): S100A9 expression was tightly associated with high grade, negative ER and PR status, high Ki67 and p53 expression, and ERBB2 and EGFR expression; it was also closely correlated to our previously described 10-protein basal signature combining IHC expression of CK5, CD10,
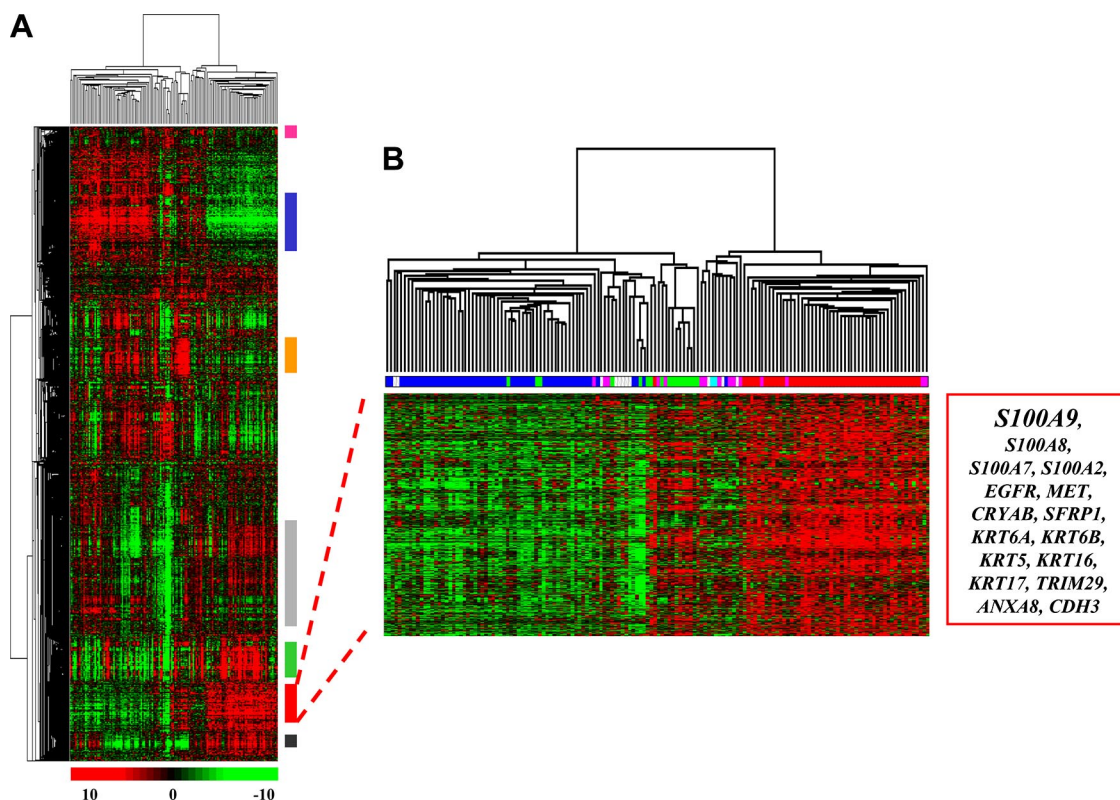
FIG. 5. **Global gene expression profiling in breast cancer tumors.** *A*, hierarchical clustering of 148 breast cancer samples and four normal breast samples and 13,743 genes/expressed sequence tags with significant variation in mRNA expression level across all samples. Each *row* represents a gene, and each *column* represents a sample. The dendrogram of samples (*above* matrix) represents overall similarities in gene expression profiles and is enlarged in *B*. *Colored bars* to the *right* indicate the locations of seven gene clusters of interest: from *top* to *bottom*, the ERBB2-related cluster (*pink bar*), the luminal/ER cluster (*blue bar*), the stromal cluster (*orange bar*), the ribosomal/metabolism cluster (*light gray bar*), the immune cluster (*green bar*), the basal cluster (*red bar*, enlarged in *B*), and the proliferation cluster (*dark gray bar*). *B*, dendrogram of samples and basal gene cluster. *Top*, dendrogram of samples: two large groups of samples are evidenced by clustering. *Middle*, the molecular subtype of samples, based on Sorlie *et al.* (2, 3) centroids, is represented according to a color ladder: *dark blue*, luminal A; *light blue*, luminal B; *red*, basal; *pink*, ERBB2-positive; *green*, normal-like; *white*, unassigned (tumors that do not fall in a particular subtype). *Bottom*, expanded view of the basal gene cluster that includes S100A9.

EGFR, CAV1, CD44, ETS1, MET, Moesin, GATA3, and CK19 (8). S100A9 expression was also associated with young age and lymph node invasion. Interestingly 73 frozen samples of this panel (which were part of the above described panel of 148 tumors) had been classified previously as basal or non-basal tumors using DNA microarrays. Again S100A9 expression was significantly associated with the basal transcriptional subtype. In addition, we evaluated the repartition of S100A9 expression across the protein-based subclasses of breast cancer we described previously from the same TMA (18) (supplemental Fig. 1). S100A9 was much more frequently expressed in tumors of cluster B (basal-like tumors, 75%) as compared with A1 (luminal A-like tumors, 25%) and A2 clusters (54%) (Fisher's exact test, $p = 8.3 \cdot 10^{-16}$). Finally S100A9 protein expression was significantly associated with reduced metastasis-free ($p = 0.007$, log-rank test) and overall survivals ($p = 0.0002$, log-rank test) (Fig. 6C). S100A9 as well as other potential prognostic markers retaining significant prognostic value in univariate

analysis such as lymph node status, ER, ERBB2, grade, vascular invasion, MIB1, and tumor size were used to build a multivariate Cox model. The final model only retained lymph node invasion, ER, grade, tumor size, and MIB1 (proliferation marker) but not S100A9 as independent prognostic factors, indicating that the negative prognosis associated with S100A9 expression might be due to its correlation with some of these parameters (Table IV). However, we also looked at the potential prognostic impact of S100A9 expression among node-negative patients. We found that a significant worse overall survival was associated with S100A9 expression in this subset of patients (hazard ratio = 4.13 (CI 95%, 1.82–9.34), $p = 0.0096$). Importantly in this subset of patients, S100A9 retained independent prognostic significance in multivariate analysis (Table IV). Thus, SELDI-based protein profiling in BCLs allowed the identification of a basal biomarker, the clinical relevance of which was emphasized by other molecular typing techniques in tumor samples.

FIG. 6. **Immunohistochemistry and S100A9 expression.** *A*, Correlation between S100A9 measurements using mass spectrometry on cytosolic protein lysates and IHC on CMA. SELDI measurements of 13300 S100A9 protein expression (normalized linear intensity) in Triton-extracted protein lysates were obtained from 24 BTCL on ProteinChip arrays and plotted against expression measured by immunohistochemistry (quick score) for the same cell lines on a CMA built as described in material and methods. Correlation coefficient (ρ) was tested for significance using Pearson's test. *B*, Expression of S100A9 protein studied by immunohistochemistry on TMA. *Left panel*, representative hematoxylin-eosin-saffron staining of a paraffin block section (25×30 mm^2) from a TMA containing 552 early breast cancer cases with 0.6-mm tumor cores. *Right panel*, immunohistochemical staining of a tumor core: low (*left*) and high (*right*) expression in representative cancer tissues. *Magnification*, $\times 400$. *C*, Kaplan-Meier analysis of the metastasis-free and overall survivals of early breast cancer patients according to S100A9 expression on TMA.
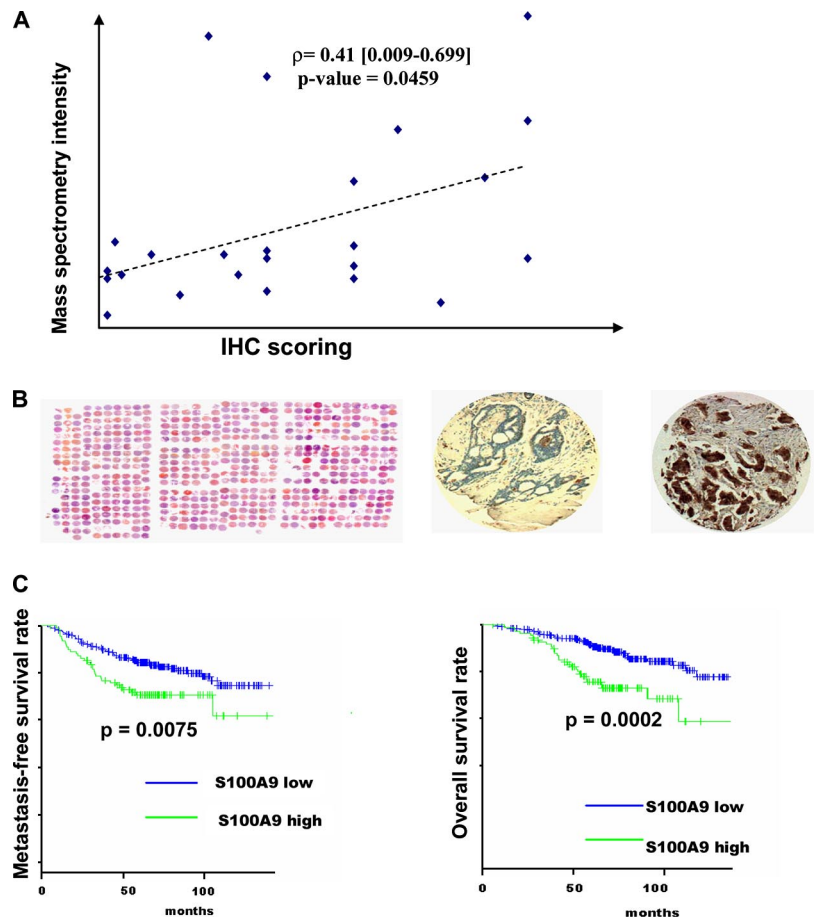
DISCUSSION

Breast cancer is a heterogeneous disease with a wide range of molecular abnormalities leading to diverse and hard to predict behavioral outcomes. Recent data emerging from large scale molecular typing technologies have allowed this molecular diversity to be deciphered and have shed light on a novel and robust classification of breast cancer, which should improve patient management (1–3, 7). Breast tumors can be classified into major subtypes on the basis of gene expression signature: luminal, ERBB2-overexpressing, and basal-like. Basal-like tumors are largely ER-, PR-, and ERBB2-negative (triple negative) and express genes characteristic of basal epithelial cells including the basal cytokeratins CK5/6 and CK17. This subtype has been associated with poor prognosis and represents a challenging issue in mammary oncology essentially for two reasons: to date, a routinely usable diagnostic biomarker of this subtype is lacking, and no specific targeted therapy is available for these cancers as opposed to endocrine therapy or ERBB2-directed approaches for luminal and ERBB2 subtypes, respectively.

Although some significant differences have been documented, breast cell lines have been shown to reflect the genomic, transcriptional, and biological heterogeneity found in primary tumors and appear to be a good system to identify many recurrent genomic and transcriptional abnormalities of primary breast tumors (22). In this study, we applied mass spectrometry-based protein profiling using SELDI-TOF MS technology to cytosolic extracts from human BCLs previously characterized for molecular subtypes by transcriptional analyses (8). Interestingly unsupervised analysis of the generated protein profiles, composed of 326 protein peaks, allowed the segregation of two phenotypic groups. These groups were different in terms of pathological and molecular features, notably ER expression but also basal-like and non-basal subtypes. This result suggests that a limited amount of protein information, extracted directly from the Triton-soluble fraction of tumor cells, can capture molecular characteristics with critical relevance to breast cancer biology. Supervised analysis generated a list of potential protein biomarkers with differential expression according to the basal or luminal phenotype. Two of them were identified by purification and a subsequent MALDI-based approach.

The first marker was a subtle post-translational modification of ubiquitin, namely the removal of two glycines at the carboxyl terminus of the protein; the quantitative level of full-length ubiquitin was apparently unchanged. Ubiquitin is a small conserved protein with 76 amino acids and is the central component of the ubiquitin-proteasome protein degradation

TABLE IV

Cox proportional hazards multivariate analyses of overall survival in overall and node-negative early breast cancer patient populations

	Variables	Hazard ratio (CI95%)	*p*-value
Overall population (n=345)	ER		
	Negative	1	
	Positive	0.52 [0.298-0.911]	0.022
	Node		
	Negative	1	
	Positive	1.84 [1.068-3.178]	0.028
	Grade		
	1-2	1	
	3	2.03 [1.163-3.544]	0.013
	MIB1		
	Low	1	
	High	2.49[1.291-4.791]	0.0065
	Tumor size		
	$\leq$20 mm	1	
	>20 mm	2.42 [1.299-4.504]	0.0054
Node negative population (n=183)	MIB1		
	Low	1	
	High	4.13 [1.55-11.01]	0.0046
	Tumor size		
	$\leq$20 mm	1	
	>20 mm	3.8. [1.371-10.559]	0.01
	S100A9		
	Low	1	
	High	2.68 [1.022-7.007]	0.045

pathway, which also involves a three-enzyme ubiquitination complex (ubiquitin-activating E1, ubiquitin-conjugating E2, and ubiquitin ligase E3), the intracellular protein ubiquitination targets, and the 26 S proteasome. Monoubiquitination occurs after attachment of a single ubiquitin to a single lysine of a substrate protein. Following monoubiquitination, a second ubiquitin molecule can be conjugated to the first one through an isopeptide bond between Gly[76] of the second ubiquitin molecule and the -NH_2 groups of one of the seven lysines (Lys[6], Lys[11], Lys[27], Lys[29], Lys[33], Lys[48], and Lys[63]) of the previously conjugated ubiquitin. After several rounds, a polyubiquitin chain may thus be conjugated to a substrate protein. Depending on the nature of the ubiquitin modification, the target protein may be destined for degradation or alternative nonproteolytic fates. For example, Gly[76]-Lys[48]-linked chains are the principal targeting signal for proteasomal degradation, whereas Lys[63]-Gly[76]-linked chains are implicated in nonproteolytic signaling such as DNA repair. Monoubiquitination regulates protein activities ranging from membrane transport to transcriptional regulation. This system regulates proteins involved in various biological processes, including the cell cycle, apoptosis, transcription, protein trafficking, signaling, DNA replication and repair, and angiogenesis (23–25). Abnormal accumulation or hyperactive degradation of these regulatory proteins may be associated with carcinogenesis. In breast cancer, a large number of abnormalities have been identified in molecules involved in protein degradation, including the p53 regulator MDM2, BRCA1, or the p27[kip1] regulator SKP2. More recently, proteins involved in the ubiquitination process as well as ubiquitin itself were shown to share a

potential prognostic value (26–31). Specifically a SELDI-based study on frozen lymph node-negative breast tumor tissues revealed a correlation between a high level of ubiquitin and good prognosis. However, it is not known whether this ubiquitin was full-length or modified at its carboxyl terminus (30). The amino acid residue 76 (Gly[76]) removed at the extreme carboxyl-terminal end of the peptide in our model is involved in every known linkage of ubiquitin to other proteins, including ubiquitin itself (32, 33). Thus, carboxyl terminus-truncated ubiquitins are virtually inactive and may act as a decoy regulating ubiquitination of specific targets and thereby their functioning in various cell processes. Carboxyl terminus-truncated ubiquitin has been found in several tissues as a result of tryptic-like protease cleavage of ubiquitin and was thought to occur during the purification process (34). Yet we have two reasons to believe that the truncated form is biologically relevant. First, we included a mixture of protease inhibitors in our cell and tissue lysates. Second, we found the truncated ubiquitin form differentially expressed between basal-like and luminal-like cells. This could mean that basal-like cells are less prone to degradation perhaps because of lower tryptic lysosomal activity.

Although the reasons for the presence of this inactive isoform in cells are unclear, its low level in basal cancer cells may be associated to a higher activity of ubiquitination in this subtype. If such a hypothesis is confirmed, it might provide a rationale to explore emerging compounds targeting protein degradation pathways, such as proteasome inhibitors, in this particular subtype with no current specific therapeutic alternatives. To further investigate the biology behind ubiquitin

truncation, we have initiated functional studies, including pull-down experiments using tumor cell lysates incubated with unmodified ubiquitin and ubiquitin with GG removed from the carboxyl terminus bound to Sepharose beads followed by MS identification of differentially bound proteins. Our preliminary results indicate that these slightly modified proteins have different binding properties, suggesting that they may have distinct signaling functions.[2] Interestingly truncated ubiquitin was also clearly identified by mass spectrometry in frozen breast cancer samples and appeared to have a significantly higher expression in luminal compared with basal samples as measured by SELDI in a small set of frozen breast cancer samples that had been characterized previously at the transcriptional level. However, the level of expression of the truncated form detected by MS in frozen whole tumors was rather low compared with the signal obtained with the pure tumor cell population from breast cancer cell lines. These differences may be due to signal dilution because of the presence of non-tumor tissue in nonmicrodissected samples (data not shown). Thus, a clear validation of truncated ubiquitin as a potential biomarker would require a larger number of samples as well as the development of alternative techniques of quantification. We are currently planning to develop specific anti-ubiquitin monoclonal antibodies that will discriminate between both ubiquitin forms and that could be used for IHC on a large TMA.

The second marker of interest, which was up-regulated in basal cancer cells, was S100A9. S100 proteins comprise a family of 23 different members characterized by sequence identity, high homology, low molecular weight, the presence of two calcium-binding EF-hands, and tissue-specific expression (35). Most of the numerous S100 genes are located in a gene cluster on 1q21, a chromosomal region prone to rearrangement during tumor development. Initially described in neutrophils and macrophages and involved in myeloid cell maturation and in inflammation, an association of S100A9 as well as other S100 protein expression with adenocarcinomas in humans has emerged. Thus, immunohistochemical investigations have shown that S100A9 protein is expressed in various epithelial tumors, including invasive breast cancer (36–40). S100A9 expression was correlated to high grade and/or poor differentiation. However, the molecular functions as well as the putative role of S100A9 in the tumor phenotype remain unclear. Recently S100A9 was shown to be involved in the metastatic processes as a chemoattractant for tumor cells that is secreted by endothelial and myeloid cells (41). Whether or not S100A9 can be secreted by tumor cells and participates in migration/invasion remains to be investigated. In our study, S100A9 expression was strongly associated with a worse prognosis, but multivariate analysis indicated that this negative impact could be due to the tight correlation between S100A9 and other pathological factors closely associated with the basal subtype such as ER negativity and high grade. However, S100A9 expression did have independent prognostic value in lymph node-negative patients, suggesting that it could be clinically useful in this increasingly important subgroup of patients. In addition, using the same MS-based approach, we recently identified another S100 protein family member associated with basal-like BCLs, namely S100A8 (m/z 10,885) (supplemental Table 4). Regarding what is known about the biology of S100 proteins and their frequent dimerization, the prognostic value of S100A9 expression might be improved by simultaneously evaluating the expression of S100A8 protein; this is currently under evaluation.

Our study has general values. First, it validates the use of MS-based techniques to subclassify breast cancer. Second, it emphasizes the great potential of proteomics in discovering new biomarkers. Third, it shows that MS-based approaches are coherent with other techniques: one marker, S100A9, also has been identified by gene and protein expression analyses. Importantly our study demonstrates, at least for pure tumor cell populations, a fair correlation between mass spectrometry intensity and IHC scoring, further validating the reliability of the technology. Fourth, it shows that an MS-based strategy may identify markers that cannot be studied by the other techniques: the ubiquitin isoform was virtually inaccessible to other tumor typing techniques including DNA microarrays or traditional immunohistochemistry using standard anti-ubiquitin antibodies. Thus, the MS-based strategy is both reliable and unique. SELDI-based protein profiling in oncology was recently investigated as a promising way to identify potential early diagnosis-related biomarkers using biological fluids, such as serum, plasma, or urine, as surrogate tissues inside which tumor fingerprints could be detected. Although various studies have suggested exciting perspectives in the diagnostic (12, 13, 42) or theragnostic fields (14), most of the proteins actually identified were highly abundant, nonspecific, host response-generated proteins. To get access to tumor-specific biomarkers, global MS-based protein profiling strategies were recently applied to solid tumor samples directly on frozen tissues including brain and lung tumors (43–45) and have shown promising ability to separate between clinically or pathologically relevant groups of tumor. Taken together with our results, such a strategy appears to be a promising method for molecular characterization of tumors and may reveal unsuspected biomarkers with potential diagnostic or theragnostic relevance.

[2] A. Gonçalves, S. Audebert, Y. Toiron, and J.-P. Borg, unpublished results.

REFERENCES

1. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. (2000) Molecular portraits of human breast tumours. *Nature* **406,** 747–752

2. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P., and Borresen-Dale, A. L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10869–10874

3. Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L., and Botstein, D. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 8418–8423

4. Troester, M. A., Hoadley, K. A., Sorlie, T., Herbert, B. S., Borresen-Dale, A. L., Lonning, P. E., Shay, J. W., Kaufmann, W. K., and Perou, C. M. (2004) Cell-type-specific responses to chemotherapeutics in breast cancer. *Cancer Res.* **64,** 4218–4226

5. Bertucci, F., Finetti, P., Rougemont, J., Charafe-Jauffret, E., Cervera, N., Tarpin, C., Nguyen, C., Xerri, L., Houlgatte, R., Jacquemier, J., Viens, P., and Birnbaum, D. (2005) Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. *Cancer Res.* **65,** 2170–2178

6. Rouzier, R., Perou, C. M., Symmans, W. F., Ibrahim, N., Cristofanilli, M., Anderson, K., Hess, K. R., Stec, J., Ayers, M., Wagner, P., Morandi, P., Fan, C., Rabiul, I., Ross, J. S., Hortobagyi, G. N., and Pusztai, L. (2005) Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin. Cancer Res.* **11,** 5678–5685

7. Sotiriou, C., Neo, S. Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S. B., Harris, A. L., and Liu, E. T. (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 10393–10398

8. Charafe-Jauffret, E., Ginestier, C., Monville, F., Finetti, P., Adelaide, J., Cervera, N., Fekairi, S., Xerri, L., Jacquemier, J., Birnbaum, D., and Bertucci, F. (2006) Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* **25,** 2273–2284

9. Bertucci, F., Birnbaum, D., and Goncalves, A. (2006) Proteomics of breast cancer: principles and potential clinical applications. *Mol. Cell. Proteomics* **5,** 1772–1786

10. Petricoin, E. F., and Liotta, L. A. (2004) SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. *Curr. Opin. Biotechnol.* **15,** 24–30

11. Kozak, K. R., Amneus, M. W., Pusey, S. M., Su, F., Luong, M. N., Luong, S. A., Reddy, S. T., and Farias-Eisner, R. (2003) Identification of biomarkers for ovarian cancer using strong anion-exchange ProteinChips: potential use in diagnosis and prognosis. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12343–12348

12. Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M., Mills, G. B., Simone, C., Fishman, D. A., Kohn, E. C., and Liotta, L. A. (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359,** 572–577

13. Zhang, Z., Bast, R. C., Jr., Yu, Y., Li, J., Sokoll, L. J., Rai, A. J., Rosenzweig, J. M., Cameron, B., Wang, Y. Y., Meng, X.-Y., Berchuck, A., van Haaften-Day, C., Hacker, N. F., de Bruijn, H. W. A., van der Zee, A. G. J., Jacobs, I. J., Fung, E. T., and Chan, D. W. (2004) Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **64,** 5882–5890

14. Goncalves, A., Esterni, B., Bertucci, F., Sauvan, R., Chabannon, C., Cubi-zolles, M., Bardou, V. J., Houvenaeghel, G., Jacquemier, J., Granjeaud, S., Meng, X. Y., Fung, E. T., Birnbaum, D., Maraninchi, D., Viens, P., and Borg, J. P. (2006) Postoperative serum proteomic profiles may predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy. *Oncogene* **25,** 981–989

15. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 14863–14868

16. Shevchenko, A., Wilm, M., Vorm, O., Jensen, O. N., Podtelejnikov, A. V., Neubauer, G., Mortensen, P., and Mann, M. (1996) A strategy for identifying gel-separated proteins in sequence databases by MS alone. *Biochem. Soc. Trans.* **24,** 893–896

17. Jaulin-Bastard, F., Saito, H., Le Bivic, A., Ollendorff, V., Marchetto, S., Birnbaum, D., and Borg, J.-P. (2001) The ERBB2/HER2 receptor differentially interacts with ERBIN and PICK1 PSD-95/DLG/ZO-1 domain proteins. *J. Biol. Chem.* **276,** 15256–15263

18. Jacquemier, J., Ginestier, C., Rougemont, J., Bardou, V. J., Charafe-Jauffret, E., Geneix, J., Adelaide, J., Koki, A., Houvenaeghel, G., Hassoun, J., Maraninchi, D., Viens, P., Birnbaum, D., and Bertucci, F. (2005) Protein expression profiling identifies subclasses of breast cancer and predicts prognosis. *Cancer Res.* **65,** 767–779

19. Kaplan, E. L., and Meier, P. (1958) Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53,** 457–481

20. Ross, D. T., and Perou, C. M. (2001) A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines. *Dis. Markers* **17,** 99–109

21. Bertucci, F., Finetti, P., Cervera, N., Charafe-Jauffret, E., Mamessier, E., Adelaide, J., Debono, S., Houvenaeghel, G., Maraninchi, D., Viens, P., Charpin, C., Jacquemier, J., and Birnbaum, D. (2006) Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers. *Cancer Res.* **66,** 4636–4644

22. Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J. P., Tong, F., Speed, T., Spellman, P. T., DeVries, S., Lapuk, A., Wang, N. J., Kuo, W. L., Stilwell, J. L., Pinkel, D., Albertson, D. G., Waldman, F. M., McCormick, F., Dickson, R. B., Johnson, M. D., Lippman, M., Ethier, S., Gazdar, A., and Gray, J. W. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10,** 515–527

23. Mani, A., and Gelmann, E. P. (2005) The ubiquitin-proteasome pathway and its role in cancer. *J. Clin. Oncol.* **23,** 4776–4789

24. Chen, C., Seth, A. K., and Aplin, A. E. (2006) Genetic and expression aberrations of E3 ubiquitin ligases in human breast cancer. *Mol. Cancer Res.* **4,** 695–707

25. Hoeller, D., Hecker, C. M., and Dikic, I. (2006) Ubiquitin and ubiquitin-like proteins in cancer pathogenesis. *Nat. Rev. Cancer* **6,** 776–788

26. Iwaya, K., Ogawa, H., Mukai, Y., Iwamatsu, A., and Mukai, K. (2003) Ubiquitin-immunoreactive degradation products of cytokeratin 8/18 correlate with aggressive breast cancer. *Cancer Sci.* **94,** 864–870

27. Wu, J. Y., Vlastos, A. T., Pelte, M. F., Caligo, M. A., Bianco, A., Krause, K. H., Laurent, G. J., and Irminger-Finger, I. (2006) Aberrant expression of *BARD1* in breast and ovarian cancers with poor prognosis. *Int. J. Cancer* **118,** 1215–1226

28. Slotky, M., Shapira, M., Ben-Izhak, O., Linn, S., Futerman, B., Tsalic, M., and Hershko, D. (2005) The expression of the ubiquitin ligase subunit Cks1 in human breast cancer. *Breast Cancer Res.* **7,** R737–R744

29. Burger, A. M., Gao, Y., Amemiya, Y., Kahn, H. J., Kitching, R., Yang, Y., Sun, P., Narod, S. A., Hanna, W. M., and Seth, A. K. (2005) A novel RING-type ubiquitin ligase breast cancer-associated gene 2 correlates with outcome in invasive breast cancer. *Cancer Res.* **65,** 10401–10412

30. Ricolleau, G., Charbonnel, C., Lode, L., Loussouarn, D., Joalland, M. P., Bogumil, R., Jourdain, S., Minvielle, S., Campone, M., Deporte-Fety, R., Campion, L., and Jezequel, P. (2006) Surface-enhanced laser desorption/ionization time of flight mass spectrometry protein profiling identifies ubiquitin and ferritin light chain as prognostic biomarkers in node-negative breast cancer tumors. *Proteomics* **6,** 1963–1975

31. Miyoshi, Y., Nakayama, S., Torikoshi, Y., Tanaka, S., Ishihara, H., Taguchi, T., Tamaki, Y., and Noguchi, S. (2006) High expression of ubiquitin carboxy-terminal hydrolase-L1 and -L3 mRNA predicts early recurrence in patients with invasive breast cancer. *Cancer Sci.* **97,** 523–529

32. Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., and Cook, W. J. (1985) Three-dimensional structure of ubiquitin at 2.8 Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 3582–3585

33. Reyes-Turcu, F. E., Horton, J. R., Mullally, J. E., Heroux, A., Cheng, X., and Wilkinson, K. D. (2006) The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell* **124,** 1197–1208

34. Wilkinson, K. D., and Audhya, T. K. (1981) Stimulation of ATP-dependent proteolysis requires ubiquitin with the COOH-terminal sequence Arg-Gly-Gly. *J. Biol. Chem.* **256,** 9235–9241

35. Gebhardta, C., Németha, J., Angel, P., and Hessa, J. (2006) S100A8 and S100A9 in inflammation and cancer. *Biochem. Pharmacol.* **72,** 1622–1631

36. El-Rifai, W., Moskaluk, C. A., Abdrabbo, M. K., Harper, J., Yoshida, C., Riggins, G. J., Frierson, H. F., Jr., and Powell, S. M. (2002) Gastric cancers overexpress S100A calcium-binding proteins. *Cancer Res.* **62,** 6823–6826

37. Arai, K., Teratani, T., Nozawa, R., and Yamada, T. (2001) Immunohistochemical investigation of S100A9 expression in pulmonary adenocarcinoma: S100A9 expression is associated with tumor differentiation. *Oncol. Rep.* **8,** 591–596

38. Arai, K., Yamada, T., and Nozawa, R. (2000) Immunohistochemical investigation of migration inhibitory factor-related protein (MRP)-14 expression in hepatocellular carcinoma. *Med. Oncol.* **17,** 183–188

39. Ott, H. W., Lindner, H., Sarg, B., Mueller-Holzner, E., Abendstein, B., Bergant, A., Fessler, S., Schwaerzler, P., Zeimet, A., Marth, C., and Illmensee, K. (2003) Calgranulins in cystic fluid and serum from patients with ovarian carcinomas. *Cancer Res.* **63,** 7507–7514

40. Cross, S. S., Hamdy, F. C., Deloulme, J. C., and Rehman, I. (2005) Expression of S100 proteins in normal human tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11 are all overexpressed in common cancers. *Histopathology* **46,** 256–269

41. Hiratsuka, S., Watanabe, A., Aburatani, H., and Maru, Y. (2006) Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. *Nat. Cell Biol.* **8,** 1369–1375

42. Petricoin, E. F., III, Ornstein, D. K., Paweletz, C. P., Ardekani, A., Hackett, P. S., Hitt, B. A., Velassco, A., Trucco, C., Wiegand, L., Wood, K., Simone, C. B., Levine, P. J., Linehan, W. M., Emmert-Buck, M. R., Steinberg, S. M., Kohn, E. C., and Liotta, L. A. (2002) Serum proteomic patterns for detection of prostate cancer. *J. Natl. Cancer Inst.* **94,** 1576–1578

43. Yanagisawa, K., Shyr, Y., Xu, B. J., Massion, P. P., Larsen, P. H., White, B. C., Roberts, J. R., Edgerton, M., Gonzalez, A., Nadaf, S., Moore, J. H., Caprioli, R. M., and Carbone, D. P. (2003) Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* **362,** 433–439

44. Schwartz, S. A., Weil, R. J., Johnson, M. D., Toms, S. A., and Caprioli, R. M. (2004) Protein profiling in brain tumors using mass spectrometry: feasibility of a new technique for the analysis of protein expression. *Clin. Cancer Res.* **10,** 981–987

45. Schwartz, S. A., Weil, R. J., Thompson, R. C., Shyr, Y., Moore, J. H., Toms, S. A., Johnson, M. D., and Caprioli, R. M. (2005) Proteomic-based prognosis of brain tumor patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry. *Cancer Res.* **65,** 7674–7681